

➤ BIPAA, un exemple d'application d'un système automatique, modulaire et FAIR de gestion de données génomiques

Stéphanie Robin, Matéo Boudet, Fabrice Legeai,
Anthony Bretaudeau



➤ Plateforme de bioinformatique



Génomique des insectes ravageurs de cultures
Département INRAE SPE (Santé des Plantes et Environnement)



Institut de Génétique,
Environnement et
Protection des Plantes,
(IGEPP) Le Rheu



Centre Inria de
l'Université de Rennes,
IRISA, Rennes



INRAE

BIPAA

14.09.23 / Journées PEPI IBIS / Stéphanie ROBIN

➤ Plateforme de bioinformatique



Nos missions :

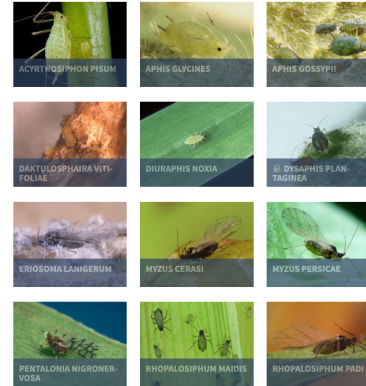
- Support et participation aux activités de recherche nécessitant de la bioinformatique : **Analyses et développement de protocoles automatisés**
- Intégration, structuration des données, assurer leur accès et leur diffusion : **Développement de systèmes d'information, infrastructures et services**
- **Formation** des biologistes aux pratiques numériques



➤ Les défis de la gestion et de la mise à disposition de données génomiques

Contexte :

- Augmentation constante du **volume**
- **Diversité** des sources
- **Hétérogénéité** des formats

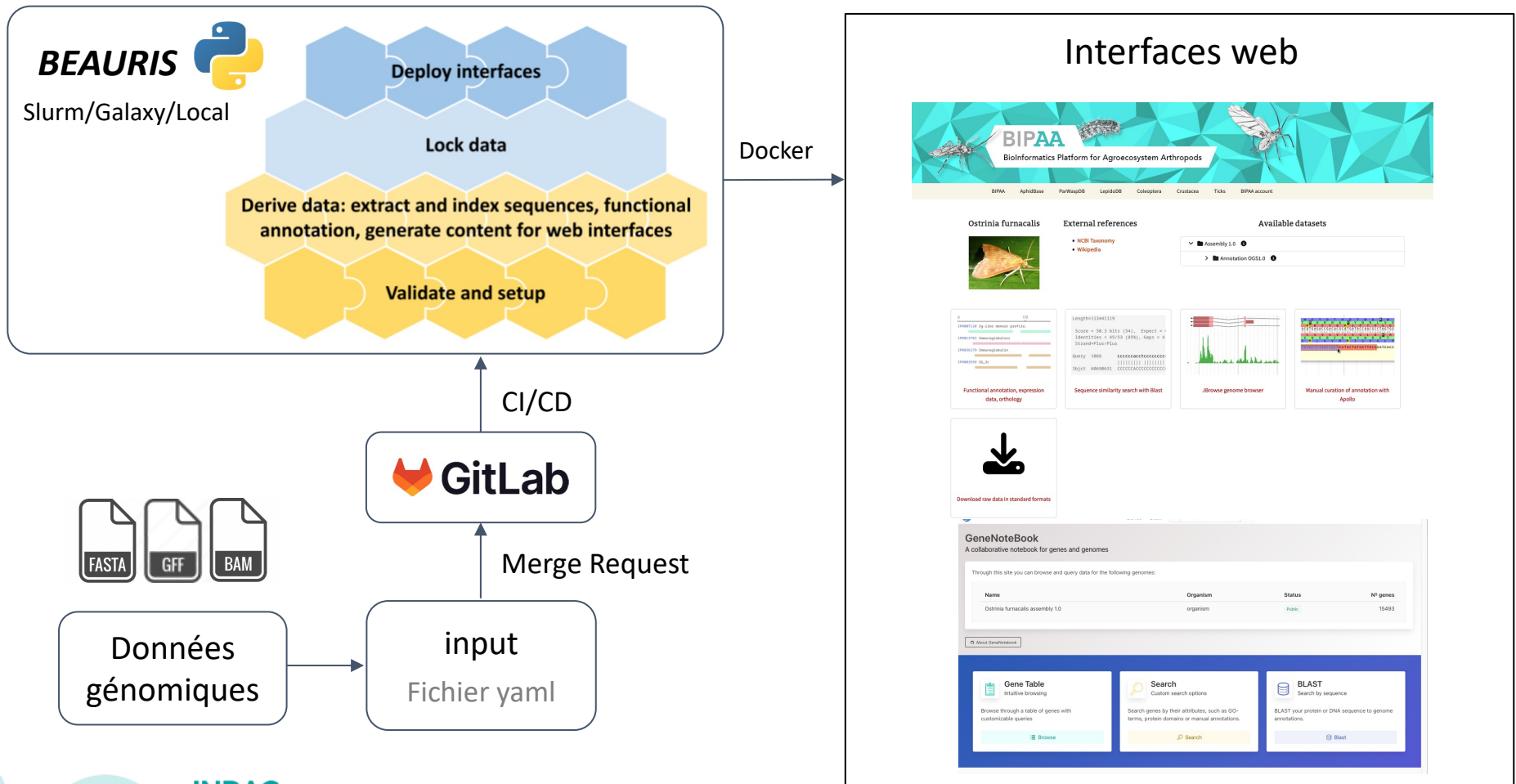


Besoins :

- **Ajout de l'information** par des pipelines automatisés (annotation fonctionnelle)
- **Intégration** de données
- **Visualisation**
- Mise à disposition d'outils pour l'**exploration**
- Déploiement de **plusieurs portails** de génomes

➤ Système d'information : fonctionnement

- Permet un **déploiement automatique, modulaire et FAIR** des données
- Utilise **Gitlab CI/CD** combiné à une **librairie Python dédiée**



➤ Métadonnées

Un fichier yaml par génome

soumis par l'utilisateur via une Merge Request

Add Hyposoter didymator

Merged ROBIN Stephanie requested to merge slittoralis into master 5 months ago

Description
du génome
et de l'espèce

```
genus: Hyposoter
species: didymator
common_name: Hyposoter didymator
restricted_to: didydb
xrefs:
  ncbi_taxonomy: 260305
  wikipedia: Hyposoter
picture:
  author: Picture copyright 2011 J.K. Lindsey
  source_url: http://www.commanster.eu/commanster/Insects/Bees/ABees/Hyposoter.didymator.html
```

Description
des fichiers
d'entrée

```
assemblies:
- version: "2.0"
  file:
    path: /home/foobar/assembly/v2.0/Hd_genome_2.0.fa
    type: fasta
  date: 2022-09-22
  source: BIPAA
  description: |
    Rescaffolding of assembly 1.0
annotations:
- version: "OGS2.0"
  file:
    path: /home/foobar/annotation/v2.0/OGS2.0_20230316.gff3
    type: gff
  date: 2023-03-16
  source: BIPAA
  description: |
    Helixer annotation
tracks:
- name: Transcriptome of Calyx-Oviduct at stage 1
  type: rnaseq
  file:
    path: /home/foobar/rnaseq/genome_v2.0/1_calyx_oviduct.markdup.sorted.bam
  type: bam
```

INRAE

BIPAA

14.09.23 / Journées PEPI IBIS / Stéphanie ROBIN

➤ Système d'intégration continue



Merge Request ouverte => environnement de **test**

Interfaces accessibles aux administrateurs

Merge Request acceptée => environnement de **production**

Données verrouillées

Interfaces publiquement accessibles (ou restreintes à un groupe d'utilisateurs)



➤ Librairie Python dédiée



Status	Pipeline	Triggerer	Stages
running In progress	need diamond to run blast2go #496164 4 -> bc1d3bf9 latest merge request		

Workflow composé de 4 grandes étapes

Vérification

- check-env check
- lint-yml check
- py-lint check
- py-test check

- setup_workdir check

- fasta-validate check
- gff-validate check

Dérivation

- gffread derive_data
- ibrowse derive_data

- func_annot derive_data

Verrouillage


- data_lock manage_data

Déploiement interfaces

- deploy_staging deploy




➤ Page d'accueil



BIPAA
Bioinformatics Platform for Agroecosystem Arthropods

BIPAA AphidBase ParWaspDB LepidoDB Coleoptera Crustacea Ticks BIPAA account

Ostrinia furnacalis

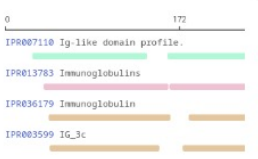


External references

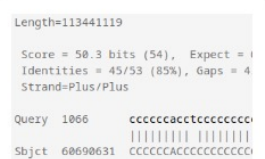
- NCBI Taxonomy
- Wikipedia

Available datasets

- Assembly 1.0
- Annotation OGS1.0




Functional annotation, expression data, orthology




Length=113441119
Score = 50.3 bits (54), Expect = 1
Identities = 45/53 (85%), Gaps = 4
Strand=Plus/Plus
Query 1866 cccccacccccccccc
 ||||| |||
Sbjct 68690631 cccccacccccccccc


Sequence similarity search with Blast



JBrowse genome browser



Manual curation of annotation with Apollo



Download raw data in standard formats



➤ Visualisation et curation

JBrowse, navigateur de génome

Available Tracks

- filter tracks
- Annotation 1
 - OGS1.0
- Reference sequence 1
 - Reference sequence

Genome Track View Help Tools

0 20,000,000 40,000,000 60,000,000 80,000,000 100,000,000 120,000,000 140,000,000

ihexagonus_scaffolds_1 ihexagonus_scaffolds_1:52619851..52667800 (47) Go

52,625,000 52,637,500 52,650,000 52,662,500

Reference sequence Zoom in to see sequence Zoom in to see sequence Zoom in to see sequence Zoom in

OGS1.0

IhexT00000421-RA

Apollo, annotations manuelles

Ostrinia scapularis 2.0 File View Help Tools

0 5,000,000 10,000,000 15,000,000 20,000,000 25,000,000

ptg000010l ptg000010l:1..28341256 (28.34 Mb) Go

User-created Annotations

- OscaOR6_2R
 - MP001087-00001
 - MP000725-00001
 - MP001089-00001
 - MP001088-00001
 - MP000791-00001
 - OscaOR3-Bis

Ostrinia scapularis (Ostrinia scapi) srobin@bipaa

Annotations Tracks Ref Sequence Search Organism Users Groups Admin

Show All Show Visible Only

Annotation Name ID All Types GO GP Prov

Rows 25 1-46 of 46

Name	Seq	Type	Length	Updated
OscaOR3-Bis	ptg000010l	gene	12,193	Sep 08, 2023
MP000791	ptg000010l	gene	4,799	Sep 08, 2023
MP001088	ptg000010l	gene	5,482	Sep 08, 2023
MP001089	ptg000010l	gene	5,522	Sep 08, 2023
MP000725	ptg000010l	gene	6,968	Sep 08, 2023
MP001087	ptg000010l	gene	5,160	Sep 08, 2023

➤ Exploration et outils

GeneNoteBook

Genes Blast 🔍

Select: Genomes Columns N° query results 15493 [Download](#)

Gene ID	Gene model	
gene-LOC114349543		<input type="checkbox"/>
gene-LOC114349544		<input type="checkbox"/>
gene-LOC114349545		<input type="checkbox"/>
gene-LOC114349546		<input type="checkbox"/>
gene-LOC114349548		<input type="checkbox"/>
gene-LOC114349549		<input type="checkbox"/>
gene-LOC114349550		<input type="checkbox"/>

BLAST

Job title

▶ Program selection

PROGRAM	<input type="text" value="blastn (nucleotide vs nucleotide)"/>
SEARCH TYPE	<input type="text" value="Somewhat similar sequences (blastn)"/>

▶ Query sequence

PASTE YOUR SEQUENCE

Index of /download/ostrinia_furnacalis/

../	16-Jun-2023 09:20
blast2go_annot/	16-Jun-2023 09:20
blast2go_gaf/	16-Jun-2023 09:20
blast2go_pdf/	16-Jun-2023 09:20
cds_fa/	16-Jun-2023 09:20
diamond/	16-Jun-2023 09:20
eggnog/	16-Jun-2023 09:20
fixed_gff/	16-Jun-2023 09:20
func_annot_readme/	16-Jun-2023 09:20
gff/	16-Jun-2023 09:20
interproscan/	16-Jun-2023 09:20
proteins_fa/	16-Jun-2023 09:20
transcripts_fa/	16-Jun-2023 09:20

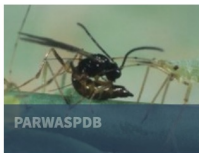
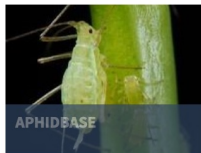
➤ Déploiement sur 2 portails



BIPAA

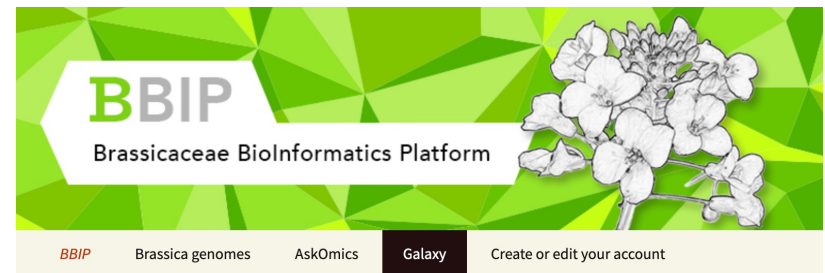
Bioinformatics Platform for Agroecosystem Arthropods

At BIPAA we provide genomic resources for multiple insect (and arthropod) species:



Also check out resources for [Coleoptera](#) and [Crustacea](#)

52 espèces



BBIP

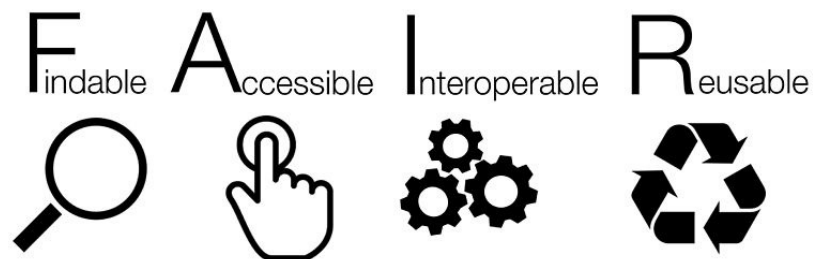
Brassicaceae Bioinformatics Platform

BBIP is a central platform to assist genomics and post-genomics programs developed on Brassicaceae by **INRAE**. It shares resources and standards critical to guarantee data integration and interoperability, and improves the coherence of the protocols set up by its various partners. Apart from giving access to the data, the main utility of BBIP is to settle an environment allowing a larger community to elaborate complex genomics analyses, browsing, mixing or crossing heterogeneous data.

The platform aims at developing particular treatments on a large panel of “omics” (e.g. genomics, transcriptomics, epigenomics, methylomics or proteogenomics) in strong collaboration with the **GenScale** team headed by Pierre Peterlongo, and **Dyliss** team headed by Olivier Dameron, both in INRIA/Irisa center in Rennes, France.

5 espèces

> Conclusion



- Système **automatisé**
- Données **structurées**
- **Modulaire, adaptable** et **déployable** n'importe où
- Analyses et déploiements **reproductibles**
- Code open source :
 - <https://gitlab.com/beaur1s/beauris>
 - <https://beauris.readthedocs.io>

> Perspectives

- Portabilité sur **d'autres portails** de génomes
- **Nouveaux types de données** : variants génomiques, orthologie, synténie, données phénotypiques
- **Nouvelles interfaces web** : JBrowse2, AskOmics, Panache, ...
- Utiliser des **standards** de données : Schema.org, RO-Crate, ...
- **Soumission** aux entrepôts de données publiques : ENA, NCBI, recherche.data.gouv.fr, ...
- **Intégration automatique** des génomes de banques publiques



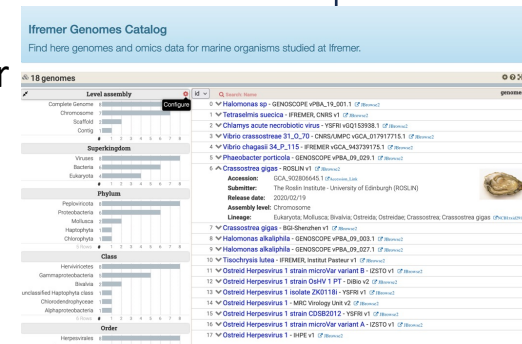
➤ Contributeur·rice·s

BIPAA / GenOuest, Rennes

- Anthony Bretaudeau
- Matéo Boudet
- Ludovic Fourteau
- Stéphanie Robin
- Fabrice Legeai

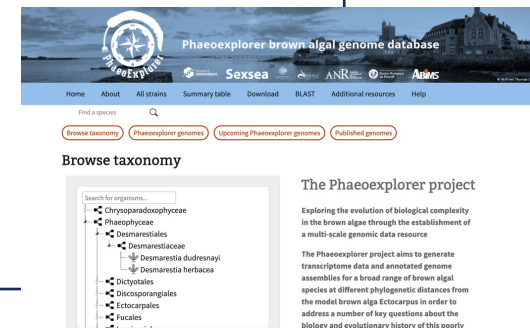
SeBiMER, IFREMER, Plouzané

- Alizée Bardon
- Alexandre Cormier
- Patrick Durand
- Laura Leroi
- Cyril Noël



ABiMS, CNRS, Roscoff

- Loraine Brillet-Guéguen
- Erwan Corre
- Arthur Le Bars
- Karine Massau



INRAE

BIPAA

14.09.23 / Journées PEPI IBIS / Stéphanie ROBIN